УДК 591.2/.6(082)

В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий, Р. Тати, В.К. Фисенко
Объединенный институт проблем информатики НАН Беларуси, г. Минск
Беларусь, 220012, г. Минск, ул. Сурганова, 6

Метод быстрого прототипирования систем нечеткого вывода при неизвестном числе классов

U.V. Anishchanka, D.A. Viattchenin, A.V. Damaratski, R. Tati, V.K. Fisenko
United Institute of Informatics Problems NAS of Belarus, c. Minsk
Belarus, 220012, c. Minsk, Surganov st., 6

The Method of Rapid Prototyping Fuzzy Inference Systems for Unknown Number of Classes

В.В. Анищенко, Д.А. Вятченін, А.В. Доморацький, Р. Тати, В.К. Фісенко
Об'єднаний інститут проблем інформатики НАН Білорусі, м. Мінськ
Білорусь, 220012, м. Мінськ, вул. Сурганова, 6

Метод швидкого прототипування систем нечіткого виводу при невідомому числі класів

В представленной статье предложена модификация метода быстрого прототипирования систем нечеткого вывода на основе результатов обработки обучающей выборки эвристическим алгоритмом возможностной кластеризации для случая априори неизвестного числа классов.
Ключевые слова: кластеризация, распределение по нечетким кластерам, нечеткое правило.

A modification of the method of rapid prototyping fuzzy inference systems based on the results of the processing of the training data set by a heuristic algorithm of possibilistic clustering for the case of a priori unknown number of classes is proposed in the presented paper.
Key Words: clustering, allotment among fuzzy clusters, fuzzy rule.

У запропонованій статті запропонована модифікація методу швидкого прототипування систем нечіткого виводу на основі результатів обробки навчальної вибірки евристичним алгоритмом можливостей кластеризації для випадку апріорі невідомого числа класів.
Ключові слова: кластеризація, розподіл за нечіткими кластерами, нечітке правило.

Введение

В задачах нечеткого управления, нечеткой классификации и принятия решений в нечеткой среде центральное место отводится системам нечеткого вывода, представляющим собой алгоритм получения нечетких заключений на основе нечетких предпосылок [1]. Главным элементом систем нечеткого вывода является база нечетких продукционных правил, в наиболее общем случае имеющих вид

l: ЕСЛИ $\hat{x}^1$ есть B_l^1 И … И $\hat{x}^m$ есть B_l^m ТО y_1 есть C_1^l И … И y_c есть C_c^l, (1)

где $l \in \{1,\ldots,c\}$ – номер правила, $\hat{x}^t \in \hat{X}^t$, $t \in \{1,\ldots,m\}$ – входные переменные, и $\hat{X}^t$ – область определения соответствующей переменной, $y_l \in Y_l$, $l \in \{1,\ldots,c\}$ – нечеткие выходные переменные, причем Y_l – область определения соответствующего заключения, а B_l^t, C_l^l – нечеткие множества с функциями принадлежности $\gamma_{B_l^t}(\hat{x}^t)$ и

$\gamma_{C_l^l}(y_l)$, определенные на соответствующих универсумах. Следует отметить, что в правиле вида (1) m посылкам соответствует c заключений, так что структура правил вида (1) в специальной литературе именуется MIMO-структурой – от англоязычного термина Multi Inputs – Multi Outputs [1], и является наиболее общей. Вместе с тем, любое правило вида (1) может быть представлено c правилами вида

$$l: \text{ЕСЛИ } \hat{x}^1 \text{ есть } B_l^1 \text{ И} \ldots \text{И } \hat{x}^m \text{ есть } B_l^m \text{ ТО } y_l \text{ есть } C_l^l, \quad (2)$$

структура которых именуется, от аббревиатуры выражения Multi Inputs – Single Output, MISO-структурой. Необходимо также указать, что условная часть некоторого нечеткого продукционного правила в специальной литературе называется антецедентом, а заключение – консеквентом [1].

На формирование базы правил систем нечеткого вывода часто оказывает влияние ряд факторов, определяемых спецификой решаемой задачи или используемого алгоритма нечеткого вывода, наиболее известными из которых являются [1]:

– Алгоритм Мамдани;
– Алгоритм Цукамото;
– Алгоритм Ларсена;
– Алгоритм Такаги – Сугэно.

База нечетких правил может формироваться, с одной стороны, экспертным путем, а с другой – на основе обработки данных обучающей выборки, для чего чаще всего используются оптимизационные методы нечеткой или возможностной кластеризации [2]. При этом исследуемая совокупность $X = \{x_1, \ldots, x_n\}$ объектов обучающей выборки обрабатывается каким-либо методом нечеткой или возможностной кластеризации с последующим проецированием значений принадлежности u_{li}, или значений типичности μ_{li} того или иного нечеткого кластера A^l, $l = 1, \ldots, c$ на координатные оси признакового пространства $I^m(X)$.

Недостатком подавляющего большинства предложенных подходов к генерированию нечетких правил на основе результатов кластеризации обучающих данных $X = \{x_1, \ldots, x_n\}$ является необходимость априорного задания числа нечетких кластеров c в искомой структуре

Целью предпринятого исследования является обобщение предложенного в [3] подхода к генерированию систем нечеткого вывода типа Мамдани, основанного на обработке обучающих данных эвристическим алгоритмом возможностной кластеризации на случай априори неизвестного числа c нечетких α-кластеров в строящемся распределении $R_c^*(X)$ по нечетким α-кластерам.

Общая схема построения нечетких классификаторов

Эвристический подход к возможностной кластеризации предложен в работе [3] и заключается нахождении так называемого распределения $R_c^*(X)$ по априори известному, или нет, числу c нечетких α-кластеров, представляющего собой частный случай возможностного разбиения. В отличие от нечеткого c-разбиения, в случае возможностного разбиения значения принадлежности μ_{li} объектов $x_i \in X$ нечетким кластерам A^l, $l = 1, \ldots, c$ интерпретируются как значения функции распределения возможностей $\pi_l(x_i)$ для нечеткого кластера A^l, $l \in \{1, \ldots, c\}$ на универсуме $X = \{x_1, \ldots, x_n\}$.

Эвристические кластер-процедуры возможностного подхода к решению задачи автоматической классификации условно делятся на два типа: реляционные, входными данными для которых является матрица нечеткого отношения сходства T, представляющая собой частный случай матрицы вида «объект-объект», и процедуры, основанные на вычислении прототипов нечетких кластеров, входными данными для которых является матрица «объект-признак», а входным параметром является некоторое расстояние между нечеткими множествами. Группа реляционных эвристических возможностных кластер-процедур включает D-AFC(c)-алгоритм, строящий распределение $R_c^*(X)$ по априори заданному числу c частично разделенных нечетких α-кластеров и представляющий собой основу подхода, D-AFC-PS(c)-алгоритм, использующий аппарат частичного обучения, и D-PAFC-алгоритм, строящий так называемое главное распределение $R_P^*(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров. В свою очередь, группа кластер-процедур, основанных на вычислении прототипов, использует транзитивное замыкание $\breve{T}$ нечеткой толерантности T и включает D-AFC-TC-алгоритм, строящий распределение $R_c^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α-кластеров, D-PAFC-TC-алгоритм, строящий главное распределение $R_P^*(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров, и D-AFC-TC(α*)-алгоритм, строящий распределение $R_c^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α-кластеров с порогом сходства, не меньшим априори задаваемого значения $\alpha^* \in (0,1]$.

Главными достоинствами эвристических алгоритмов возможностной кластеризации по сравнению с традиционными оптимизационными нечеткими и возможностными кластер-процедурами являются:

– устойчивость результатов классификации;

– высокая точность результатов классификации при минимуме, а в ряде случаев – при полном отсутствии априорной информации, и, как следствие, минимальном наборе или отсутствии задаваемых параметров;

– содержательная ясность и простота предложенного подхода, следствием чего является простота в интерпретации результатов.

Указанные особенности эвристических возможностных кластер-процедур позволяют их использовать для решения различных задач [3]: построения множества помеченных объектов для использования с нечеткими кластер-процедурами, предложенными В. Педричем [4], классификации нечетких отношений предпочтения с целью построения единственного группового решения на множестве альтернатив, отбора наиболее информативных признаков в анализируемом наборе данных, а также для построения базы правил системы нечеткого вывода типа Мамдани, для чего было предложено использовать D-AFC(c)-алгоритм. Несмотря на высокую точность, недостатком такого подхода является необходимость априорного задания в качестве параметра числа c классов, соответствующего числу нечетких правил, либо нахождение значения c с помощью некоторой меры валидности, что существенно затрудняет использование указанного метода в реальном масштабе времени. Приемлемым решением для генерирования системы нечеткого вывода в случае априорной неизвестности числа c нечетких α-кластеров в искомом распределении $R_c^*(X)$, и, как следствие, числа нечетких правил, является использование эвристических алгоритмов возможностной кластеризации, основанных на вычислении прототипов нечетких α-кластеров.

Без нарушения общности, приведенная ниже схема излагается для случая использования D-AFC-TC-алгоритма:

1. Исходные данные о совокупности объектов $X = \{x_1, \ldots, x_n\}$, представленные в форме матрицы «объект-признак», обрабатываются D-AFC-TC-алгоритмом с использованием выбранного расстояния между нечеткими множествами $d(x_i, x_j)$, $i, j = 1, \ldots, n$, результатом чего является распределение $R_c^*(X)$ по неизвестному c числу полностью разделенных нечетких α-кластеров, значение порога сходства $\alpha \in (0,1]$, а также координаты прототипов $\overline{\tau}^l$, $l = 1, \ldots, c$ нечетких α-кластеров $A_{(\alpha)}^l$, $l = 1, \ldots, c$;

2. На основе исходных данных и результатов кластеризации генерируется система нечеткого вывода, содержащая c нечетких правил;

3. На основе интерпретации результатов кластеризации, консеквентам нечетких правил присваиваются метки соответствующих классов.

Предложенный подход к генерированию системы нечеткого вывода целесообразно проиллюстрировать на примере в сравнении с результатами, полученными с помощью D-AFC(c)-алгоритма.

Иллюстративный пример

Для демонстрации преимущества указанного подхода целесообразно обратиться к простому набору данных, для чего были выбраны двумерные данные П. Снита и Р. Сокала [5], изображенные на рис. 1.

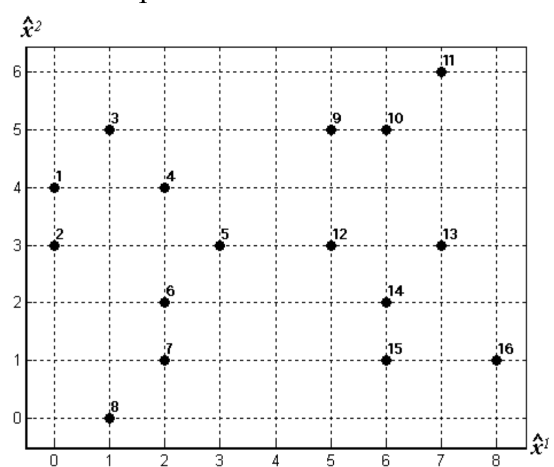


Рисунок 1 – Данные для проведения вычислительного эксперимента

При проведении вычислительного эксперимента с использованием D-AFC(c)-алгоритма и квадрата относительного Евклидова расстояния [6]

$$d(x_i, x_j) = \frac{1}{m} \sum_{t=1}^{m} \left(\mu_{x_i}(x^t) - \mu_{x_j}(x^t) \right)^2, \; i, j = 1, \ldots, n, \tag{3}$$

где $\mu_{x_i}(x^t)$ – значение принадлежности объекта $x_i \in X$ соответствующему нечеткому множеству, определенному на универсуме признаков $X^t = \{x^1, \ldots, x^m\}$, число классов полагалось равным двум, $c = 2$, в результате чего было получено распределение по частично разделенным нечетким α-кластерам при значении порога сходства $\alpha = 0.8194$. Значения принадлежности объектов нечетким α-кластерам полученного

распределения $R_c^*(X)$ представлены на рис. 2, где значения принадлежности объектов первому кластеру обозначены символом ○, а второму – соответственно, символом ■.

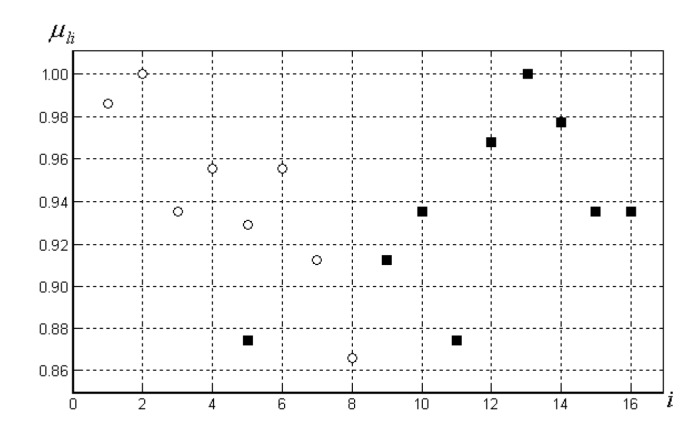


Рисунок 2 – Значения принадлежностей объектов классам
при использовании D-AFC(c)-алгоритма

Объект x_2 является типичной точкой нечеткого α-кластера $A^1_{(\alpha=0.8194)} \in R^*_{c=2}(X)$, а объект x_{13} – типичной точкой нечеткого α-кластера $A^2_{(\alpha=0.8194)} \in R^*_{c=2}(X)$. Объект x_5 оказался принадлежащим обоим нечетким α-кластерам, однако $\mu_{15} > \mu_{25}$. Сгенерированная система нечеткого вывода представлена на рис. 3, где изображен пример классификации объекта x_5.

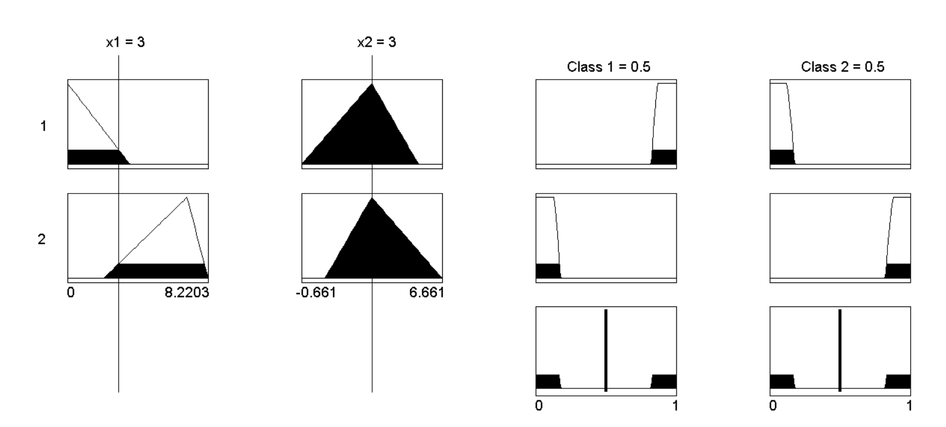


Рисунок 3 – Классификация объекта исследуемой совокупности
системой нечеткого вывода, построенной с помощью D-AFC(c)-алгоритма

Классифицируемый объект оказывается принадлежащим обоим классам в равной степени, что может трактоваться как неопределенность классификации.

В свою очередь, при проведении эксперимента с использованием D-AFC-TC-алгоритма и квадрата относительного Евклидова расстояния (3), было получено распределение по двум полностью разделенным нечетким α-кластерам при значении порога сходства $\alpha = 0.9687$, где типичной точкой нечеткого α-кластера $A^1_{(\alpha=0.9687)} \in R^*_{c=2}(X)$ является объект x_6, а типичной точкой нечеткого α-кластера $A^2_{(\alpha=0.9687)} \in R^*_{c=2}(X)$ – объект x_{10}. Значения принадлежности объектов нечетким α-кластерам полученного распреде-

ления $R_c^*(X)$, представлены на рис. 4, где, как и выше, символом ○ обозначены значения принадлежности объектов первому кластеру, а символом ■ – второму.

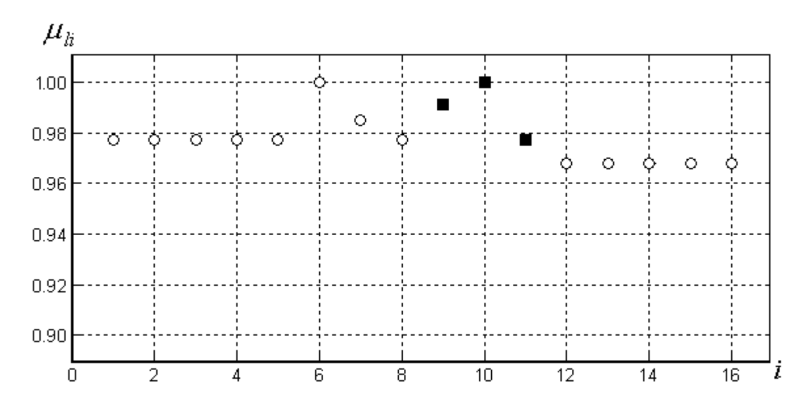


Рисунок 4 – Значения принадлежностей объектов классам при использовании D-AFC-TC-алгоритма

Очевидно, что результат, полученный с помощью D-AFC-TC-алгоритма, является менее точным, чем результат, полученный с помощью D-AFC(c)-алгоритма, что является следствием использования операции транзитивного замыкания, искажающего геометрическую структуру исследуемой совокупности. Система нечеткого вывода, сгенерированная на основе результатов, полученных с помощью D-AFC-TC-алгоритма, представлена на рис. 5.

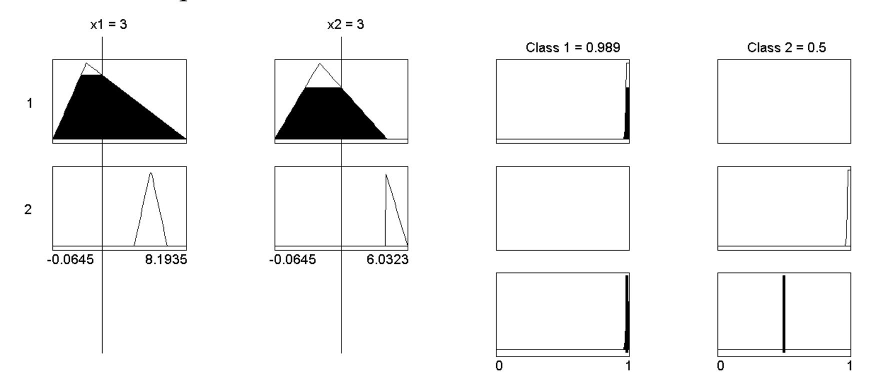


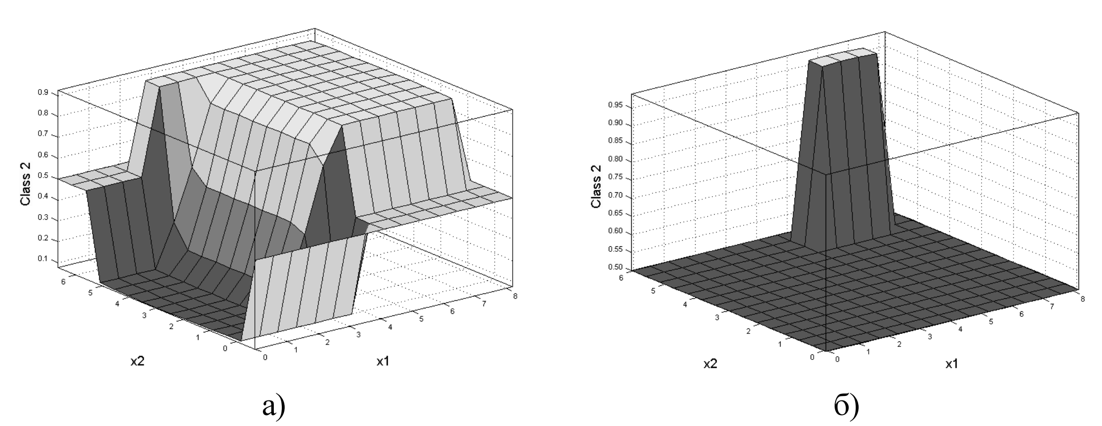
Рисунок 5 – Классификация объекта исследуемой совокупности системой нечеткого вывода, построенной с помощью D-AFC-TC-алгоритма

Классифицируемый объект x_5 оказывается однозначно принадлежащим первому нечеткому α-кластеру с высоким значением принадлежности, что соответствует результатам кластеризации, изображенным на рис. 4. Необходимо также отметить, что в консеквентах нечетких правил сгенерированной системы отсутствует функция принадлежности «низкая принадлежность объекта классу» вследствие полной разделимости нечетких α-кластеров распределения $R_c^*(X)$, генерируемого с помощью кластер-процедур, основанных на вычислении прототипов.

Различие между обеими сгенерированными системами нечеткого вывода можно также продемонстрировать на примере поверхности нечеткого вывода, визуализирующей зависимость выходных переменных от входных. На рис. 6 изображены поверхности нечеткого вывода, демонстрирующие зависимость выходной переменной Class 2 от обоих входных переменных для построенных в результате двух экспериментов систем нечеткого вывода.

Анализ поверхностей нечеткого вывода позволяет сделать заключение об адекватности сгенерированных систем соответствующим результатам кластеризации обучающих данных.

Предложенный подход позволил соответствующим образом модифицировать программный модуль FIS Generator [7], как это изображено на рис. 7.



а) б)

Рисунок 6 – Вид поверхностей нечеткого вывода для моделей, разработанных на основе результатов кластеризации обучающих данных:
а) с помощью D-AFC(c)-алгоритма; б) с помощью D-AFC-TC-алгоритма

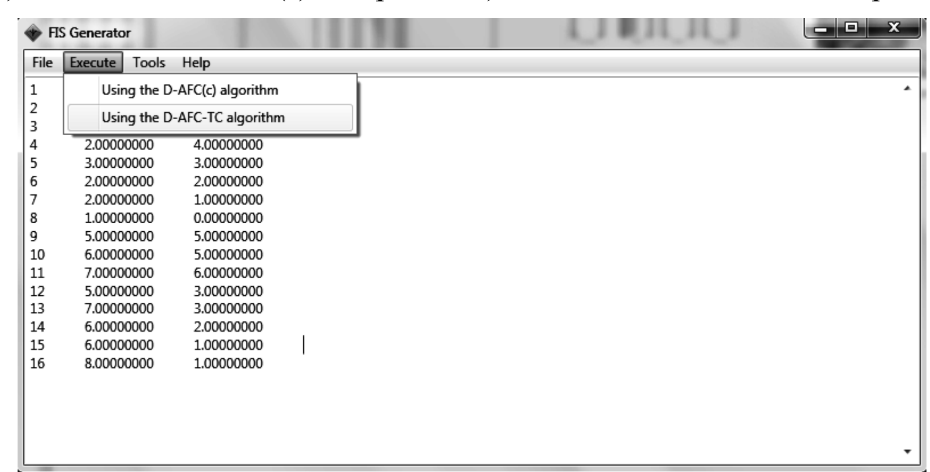


Рисунок 7 – Вид главного окна программного модуля FIS Generator
и меню выбор алгоритма кластеризации

После предварительной обработки данных D-AFC-TC-алгоритмом с ручным выбором в главном окне программы параметров кластеризации, таких, как вид используемого расстояния между нечеткими множествами [6], метод нормировки исходных данных, вид используемого критерия качества классификации и значение показателя точности вычислений, выполняется генерирование правил системы нечеткого вывода. Результат кластеризации выводится в рабочее окно программы, после чего производится непосредственное генерирование правил прямо в файл, указанный пользователем. Все вычисления выполняются в фоновом потоке и сопровождаются окном, информирующем о степени обработки данных. Генерируемая с помощью разработанного программного модуля FIS Generator система нечеткого вывода представляет собой внешний файл с расширением fis, сохраняемый пользователем с помощью стандартного диалогового окна и запускаемый с помощью среды MATLAB® версии не ниже 7.6.0 [8].

Простота реализованного метода, интуитивно понятный интерфейс, более высокая, по сравнению с D-AFC(c)-алгоритмом, скорость обработки данных D-AFC-TC-алгоритмом, а также возможность генерирования систем нечеткого вывода в режиме времени, близком к реальному, делают модифицированный программный модуль FIS Generator удобным средством для моделирования нечеткого вывода при проектировании систем различного назначения, требующими высокой скорости обработки данных.

Выводы

В статье предложено обобщение метода генерирования систем нечеткого вывода типа Мамдани для случая априорной неизвестности числа классов. Предложенное обобщение базируется на результатах обработки обучающей выборки эвристическими алгоритмами возможностной кластеризации, основанными на основанных на вычислении прототипов нечетких α-кластеров. Метод отличается простотой и содержательной осмысленностью количественной оценки результатов классификации.

Недостатком предложенного подхода является более низкая, в сравнении с D-AFC(c)-алгоритмом, точность классификации обучающих данных D-AFC-TC-алгоритмом, являющаяся следствием использования операции транзитивного замыкания, в силу чего генерируемая система нечеткого вывода является менее релевантной, чем система, генерируемая с помощью D-AFC(c)-алгоритма; в то же время, при использовании D-AFC-TC-алгоритма существенно возрастает скорость вычислений, что позволяет генерировать системы нечеткого вывода в режиме времени, близком к реальному. Кроме того, еще одним достоинством предложенного обобщения является отсутствие необходимости построения функций принадлежности «низкая принадлежность объекта классу» консеквентов нечетких правил ввиду полной разделимости нечетких α-кластеров распределения $R_c^*(X)$, являющегося результатом обработки данных D-AFC-TC-алгоритмом.

Таким образом, предложенное обобщение метода быстрого прототипирования систем нечеткого вывода целесообразно к применению в случае отсутствия априорной информации о числе классов, отсутствия необходимости очень высокой точности классификации и необходимости обработки данных в реальном масштабе времени.

Литература

1. Борисов В.В. Нечеткие модели и сети / В.В. Борисов, В.В. Круглов, А.С. Федулов. – М. : Горячая линия – Телеком, 2007. – 284 с.
2. Bezdek J.C. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing / J.C. Bezdek, J. Keller, R. Krishnapuram, N.R. Pal. – New York : Springer Science, 2005. – 776 p.
3. Viattchenin D.A. A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications / D.A. Viattchenin. – Heidelberg: Springer, 2013. – 227 p.
4. Pedrycz W. Algorithms of fuzzy clustering with partial supervision / W. Pedrycz // Pattern Recognition Letters. – 1985. – Vol. 3. – P. 13-20.
5. Sneath P.H.A. Numerical Taxonomy / P.H.A. Sneath, R.R. Sokal. – San Francisco : Freeman, 1973. – 573 p.
6. Кофман А. Введение в теорию нечетких множеств / А. Кофман; пер. В.Б. Кузьмина. – М. : Радио и связь, 1982. – 432 с.
7. Анищенко В.В. Обучение консеквентов нечетких правил, построенных на основе эвристической возможностной кластеризации / В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий // Информатика. – 2011. – № 3. – С. 98-111.
8. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. – СПб.: БХВ-Петербург, 2003. – 736 с.

Literatura

1. Borisov, V.V. Nechetkiye modeli i seti. M.: Goryachaya Liniya – Telekom. 2007. 284 S.
2. Bezdek, J.C. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. New York: Springer Science. 2005. 776 P.
3. Viattchenin, D.A. A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications. Heidelberg: Springer. 2013. 227 P.
4. Pedrycz, W. Pattern Recognition Letters. 1985. Vol. 3. P. 13-20.
5. Sneath, P.H.A. Numerical Taxonomy. San Francisco: Freeman. 1973. 573 P.
6. Kaufman, A. Vvedenie v teoriy nechetkih mnojestv. M.: Radio i Svyaz'. 1982. 432 S.
7. Anishchanka, U. Informatics. 2011. No. 3. S. 98-111.
8. Leonenkov, A.V. Nechetkoye modelirovanie v srede MATLAB i fuzzyTECH. S.-Peterburg: BHV-Peterburg. 2003. 736 S.

RESUME

U.V. Anishchanka, D.A. Viattchenin, A.V. Damaratski, R. Tati, V.K. Fisenko

The Method of Rapid Prototyping Fuzzy Inference Systems for Unknown Number of Classes

The problem of fuzzy classifiers constructing is very actual in different areas. The paper deals with the problem of rapid prototyping fuzzy inference systems for a case of unknown number of classes. For the purpose, a generalization of the method based on the heuristic approach to possibilistic clustering is proposed. The generalization based on prototype-based heuristic algorithms to possibilistic clustering for the training data processing. The algorithms are based on the transitive closure of the initial fuzzy tolerance. So, the allotment among unknown number of fully separated fuzzy clusters is the clustering result and the corresponding fuzzy inference system can be generated immediately. An illustrative example is given for Sneath and Sokal's two-dimensional data set and the result seems to be satisfactory. Some preliminary conclusions and recommendations are formulated.

Статья поступила в редакцию 09.04.2013.